PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-312058
(43)Date of publication of application : 09.11.1999

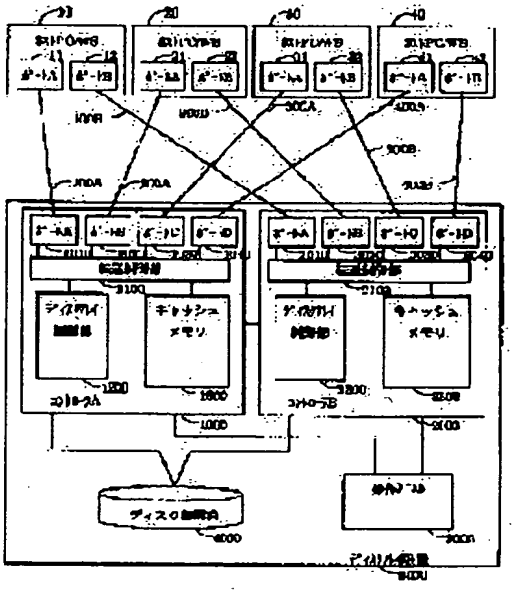
(51)Int.Cl. G06F 3/06
G06F 11/20
G06F 12/08
G06F 12/16

(21)Application number : 10-118127
(22)Date of filing : 28.04.1998

(71)Applicant : HITACHI LTD
(72)Inventor : MATSUMOTO YOSHIKO
MURAOKA KENJI
TAKAMOTO KENICHI
KOBAYASHI MASAAKI

(54) STORAGE SUB-SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To provide high performance by making it unnecessary to operate any exclusive control or the like under plural controller control in a normal time even in any connection configuration.
SOLUTION: When a disk array device 5000 in a dual controller constitution and plural port constitution is connected with plural hosts as a redundant constitution, the assignment of a logical volume is divided between controllers 1000 and 2000, and the logical volume is shared by each controller. Even when any host, bus, controller, and cache failure are caused, plural switching means on the occurrence of the failure are provided by a host connection environment, and the switching is not operated as necessary. Then even in the redundant constitution, low overhead in the same level as that in a single constitution can be realized a normal time, and any performance deterioration in the switching processing on the occurrence of each kind of failure can be prevented.



LEGAL STATUS

[Date of request for examination] 18.04.2005

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平11－312058

(43)公開日　平成11年(1999)11月９日

(51)Int.Cl.[6]		識別記号	FI		
Ｇ０６Ｆ	3/06	３０４	Ｇ０６Ｆ	3/06	３０４Ｂ
	11/20	３１０		11/20	３１０Ｆ
	12/08			12/08	Ｇ
					Ｊ
	12/16	３１０		12/16	３１０Ｊ

審査請求　未請求　請求項の数13　ＯＬ　（全８頁）

(21)出願番号　特願平10－118127

(22)出願日　平成10年(1998)４月28日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　松本　佳子
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　村岡　健司
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　高本　實一
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(74)代理人　弁理士　小川　勝男

最終頁に続く

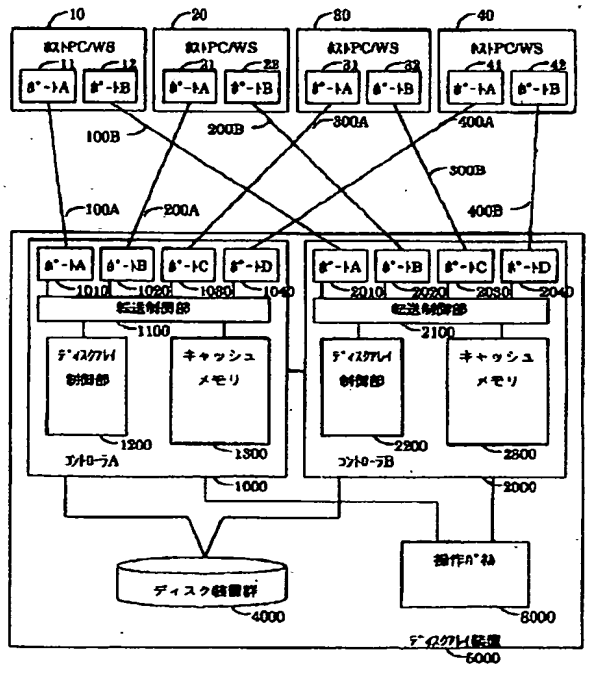
(54)【発明の名称】　記憶サブシステム

(57)【要約】

【課題】複数ホスト複数パス構成での冗長構成の制御装置との接続は、論理ボリュームを共有するしないによらず、各部分の障害時、処理の続行は可能であるが、正常時のオーバヘッドの増加により、シングル構成時との性能差が問題となっている。

【解決手段】デュアルコントローラ構成、複数ポート構成のディスクアレイ装置で複数ホストに冗長構成として接続された時、論理ボリュームの割当てをコントローラ間で分割してコントローラ毎に論理ボリュームを分担する。ホスト、パス、コントローラ、キャッシュ障害時でも、ホスト接続環境により障害時の切替え手段を複数の提供し、場合によっては切替えないようにする。そして、冗長構成においても、正常時もシングル構成と変わらない低オーバヘッドと、各種障害時の切替え処理における性能劣化を防ぐことを実現する。

図１



【特許請求の範囲】
【請求項１】記憶領域が複数の論理ボリュームに分割される記憶装置群と、
ホストコンピュータに接続され、前記ホストコンピュータと記憶装置との間を転送されるデータを一時的に保存するキャッシュメモリを有するコントローラを複数有し、前記複数のコントローラが前記記憶装置群を制御する記憶サブシステムにおいて、
前記コントローラは、処理を担当する前記論理ボリュームを割り当てられ、前記キャッシュメモリに複数の前記コントローラのための記憶領域を有することを特徴とする記憶サブシステム。
【請求項２】請求項１記載の記憶サブシステムにおいて、
前記コントローラは、複数の前記コントローラのための前記キャッシュメモリの記憶領域を、該コントローラが処理を担当する論理ボリュームのための領域に分割することを特徴とする記憶サブシステム。
【請求項３】請求項１記載の記憶サブシステムにおいて、
前記コントローラが前記ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領した時は、該コントローラは、該論理ボリュームの処理を担当していたコントローラの全ての論理ボリュームの処理を担当すること特徴とする記憶サブシステム。
【請求項４】請求項１あるいは２記載の記憶サブシステムにおいて、
前記コントローラが前記ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領した時は、該コントローラは、該論理ボリュームの処理を担当すること特徴とする記憶サブシステム。
【請求項５】請求項１記載の記憶サブシステムにおいて、
コントローラは、ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領した時は、該論理ボリュームの処理担当コントローラに処理要求内容を通信し、通信を受領したコントローラが該論理ボリュームに対する処理を行い、処理結果を要求元コントローラに通信することを特徴とする記憶サブシステム。
【請求項６】請求項１記載の記憶サブシステムにおいて、
コントローラがホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領した時の処理方法を選択するための入力手段を有することを特徴とする記憶サブシステム。
【請求項７】請求項２記載の記憶サブシステムにおいて、
コントローラ当たりのキャッシュメモリ領域の容量及び、前記論理ボリュームのための領域の容量を指定するための入力手段を有することを特徴とした制御装置を備えた記憶サブシステム。
【請求項８】請求項１記載の記憶サブシステムにおいて、
複数の前記コントローラのキャッシュメモリのそれぞれに前記複数の論理ボリュームのデータを格納することを特徴とする記憶サブシステム。
【請求項９】請求項１記載の記憶サブシステムにおいて、
複数の前記コントローラのキャッシュメモリのそれぞれに前記複数の論理ボリュームのデータを格納するか、前記コントローラのそれぞれが処理を担当する論理ボリュームのデータのみを格納するかを選択する入力手段を有することを特徴とする記憶サブシステム。
【請求項１０】請求項１記載の記憶サブシステムにおいて、
前記コントローラに障害が発生したときには、該障害コントローラで担当していた論理ボリュームの処理を正常に動作しているコントローラの担当に切り替えることを特徴とする記憶サブシステム。
【請求項１１】請求項３記載の記憶サブシステムにおいて、
コントローラがホストコンピュータから処理を担当していない論理ボリュームに対する処理要求を受領したときは、キャッシュメモリ上の管理情報のみの変更で担当の切替えを行うことを特徴とする制御装置を備えた記憶サブシステム。
【請求項１２】請求項４記載の記憶サブシステムにおいて、
コントローラがホストコンピュータから処理を担当していない論理ボリュームに対する処理要求を受領したときは、キャッシュメモリ上の切替え対象論理ボリュームのうち前記記憶装置群の記憶領域に格納されていないデータを切替え先のコントローラ内のキャッシュメモリ上にコピーして該論理ボリュームの処理を担当するコントローラを切替えることを特徴とする記憶サブシステム。
【請求項１３】請求項２記載の記憶サブシステムにおいて、
前記コントローラは、前記キャッシュメモリの論理ボリュームのための記憶領域を、該論理ボリュームの負荷に応じて変更することを特徴とした記憶サブシステム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、複数台のホスト又は複数のポートを持つ単一のホストと接続する制御装置を備えた記憶サブシステムにおいて、特に詳細には、記憶サブシステムとしての障害時の信頼性と可用性の向上技術及び障害時の交代系への処理の切替え性能の向上方式技術に関する。
【０００２】
【従来の技術】コントローラ及びディスク等の記憶装置

に冗長性を持たせたものとして、従来、特開平４－２１５１４２に記載された記憶サブシステムがある。この記憶サブシステムは、２重の系で構成され、一方の系が現用系として稼働し、他方の系が待機系として稼働している。そして、現用系のディスク装置の記憶情報を両系からアクセス可能な共用ディスク装置を介して予備系のディスク装置に複写すること、あるいは、現用系コントローラ障害時は、予備系のコントローラによって、現用系のディスク装置の記憶情報を抽出可能とすることで、コントローラ及びディスク装置障害時のデータ保全性の向上を計っていた。

【０００３】又、従来複数のコントローラで、且つ複数ポート構成で記憶媒体を共有する制御装置は基幹系オープン市場で見られるように、任意の論理ボリュームは複数のパスからアクセスが可能であり、又、全論理ボリュームがどのコントローラからも均一にアクセス可能な構成であった。

【０００４】

【発明が解決しようとする課題】コンピュータシステムの大規模化、データ処理の高速化、データの大容量化に伴い、記憶制御装置に対する高性能、高信頼性、可用性の向上が強く望まれている。

【０００５】従来技術では、制御装置側の2重系、また、キャッシュの2重化等により、制御装置内に閉じた信頼性、可用性の向上のみが行われていた。コンピュータシステム全体を考えると、制御装置の２重化とともに、ホスト、および制御装置に対するアクセスパスの２重化が必須であり、これを実現するためには複数コントローラ構成、複数ポート構成が必須となる。

【０００６】しかし、このような構成において従来の基幹系オープン市場のような、全論理ボリュームがどのコントローラからも均一にアクセス可能な構成を実現するためには、複数のコントローラ間でホストのデータを一時格納するキャッシュの管理情報を共有することが必須である。キャッシュの管理情報を共有するには、排他制御を行う為に専用のロック制御、排他情報の管理等をプロセッサで行う必要が有るため、１つのコントローラ構成での性能にくらべ、複数コントローラ構成での性能劣化の割合が大きい。よって信頼性、可用性を向上させるためにその代替として性能を犠牲にするしかなかった。

【０００７】システムとして冗長構成をとりながら、冗長構成でないシングル構成時に比べ、性能を維持するためには、どのコントローラからも、全ての論理ボリュームのアクセスが可能でありながら、コントローラ間でのロック制御、排他制御等が不必要であることが必要である。よって、通常のアクセス形態としては、コントローラ単位に論理ボリュームの担当を固定化させる方式が考えられるが、単に固定的な制御では、ホスト、ホスト－制御装置間パス、制御装置の障害時、当該担当論理ボリュームへの処理の代替えが不可能となり、冗長構成であるとは言えない。よって、通常時（正常時）は、各々担当業務を低オーバヘッドで実現し、障害時は、肩代わりができることが必要である。

【０００８】しかし、各コントローラへの接続形態として、1つのポートで、例えば、デイジーチェーン接続により複数のホストより接続されている。または、コントローラが複数のポート構成で、各々のポートから異なるホストにて接続されている。さらに、各ホストが、他コントローラ側にも交代パスとして接続しているような構成の時、１つのホスト側の障害、又はパス障害により、他系コントローラに論理ボリュームを切り替えようとした時、コントローラの担当の切替えとして、他系コントローラの担当論理ボリューム全てを、一括して切り替えてしまう方法では、それ以外のホストが通常処理のままのアクセスを行うことができなくなる。よって、複数のホストと接続する場合には、各々のホスト、パス障害にあった切替え手段が必要となる。

【０００９】又、近年フォールトトレラントの強いニーズにより、ホスト側にも、フォールトトレラント機能であるパス切替え機能がサポートされつつある。代表例としては、HPのＡｌｔｅｒｎａｔｅ Ｌｉｎｋや、Ｓａｆｅ Ｐａｔｈ（パス切替えソフト）等があげられる。この本機能では、ホスト側が交代パス接続しており、任意パスでの障害を認識すると、自動的に他系のパスへ処理を移し（論理ボリュームを移す）処理を再開する。障害の認識は、タイムアウト等によるものが多く、他系に処理を移した後の最初のＩ／Ｏで一定時間以上処理が遅延すると、さらに障害と認識される危険性が高い。よって、コントローラの担当の切替え処理には、Ｉ／Ｏを受領して認識するわけだが、その後の切替え処理時間は、障害認識されない短い時間で行うことが必須条件となる。

【００１０】

【課題を解決するための手段】前述の課題を解決する為に、各々のコントローラが、複数のパス接続ポートとデータを一時的に格納するキャッシュ機構を備え、ホストからのデータを記憶する記憶媒体を共有する複数のコントローラで構成された制御装置で、論理ボリューム（ＬＵＮ）の担当の指定、及び非担当側コントローラへのＩ／Ｏ受領時の論理ボリュームの担当の切替え方法の指定、及び論理ボリュームに対するキャッシュの割当て方法の指定、及びキャッシュへの１重書き／２重書きの指定をする手段と、担当以外の論理ボリュームへの要求時、切り替える制御、及び他系コントローラへ通信する制御を司るプロセッサと、キャッシュへ２重書き／１重書きの切替えをデータ転送を制御する手段と、キャッシュメモリをコントローラ、さらには、ＬＵＮ単位に分割して使用する手段と、ＬＵＮ毎の負荷情報、及びＬＵＮの担当により、最適な切替え方法を判断するプロセッサより実現される。

【００１１】

【発明の実施の形態】以下、本発明の１実施例を図面を用いて説明する。

【００１２】図１は、本実施例に関するディスクアレイ装置を含むシステム全体の１構成例である。図１において、１０、２０、３０、４０はホストコンピュータ、５０００はディスクアレイ装置であり、各々のホストより、ＳＣＳＩバス等で接続されている。１０００、２０００はデュアル構成をとるコントローラ部、４０００はホストからのデータを格納するディスク装置群、３０００はＬＵＮの指定等を行う操作パネルである。ディスク装置群４０００は、アレイ構成であることが多い。

【００１３】各コントローラは、バスプロトコル制御を行う４つのポート１０１０、１０２０、１０３０、１０４０、２０１０、２０２０、２０３０、２０４０を備え、ホスト側の各ポートに接続される。又、転送制御部１１００、２１００は、キャッシュメモリ１３００、２３００とホスト間のデータ転送制御を司り、又１重書き／２重書きの制御も行う。又、ディスクアレイ制御部１２００、２２００はコントローラＡ、コントローラＢ全体を制御している。

【００１４】図２にキャッシュメモリの構成を示す。キャッシュメモリ１３００、２３００は、管理情報１４００，２４００、コントローラＡ用エリア１５００、２５００、コントローラＢ用エリア１６００、２６００からなる。各々１４００と２４００、１５００と２５００、１６００と２６００は２重書きされる。キャッシュメモリ１３００、２３００の管理情報１４００、２４００は、２重書きにより同一構成であるため、コントローラＡ側のキャッシュメモリ１４００を例にとり管理情報の構成を説明する。管理情報１４００は、コントローラ間の通信情報１４１０と、コントローラ毎のＬＵＮ情報１４２０、キャッシュ内データを管理するデータ管理情報１４３０からなる。データ管理情報１４３０は、コントローラ毎に２分割されている。

【００１５】キャッシュメモリ１３００、２３００は信頼性の為、不揮発性であることが望ましい。ホストが、既に冗長構成をとっていて、例えば、ディスクアレイ装置５０００を２つ用意し、１方を現行機、他方を予備機として使用しているような場合、敢えてディスクアレイ側での冗長度を要求する必要がない場合、操作パネル３０００によるユーザからの指示により、キャッシュを１重書きとして使用する場合は、コントローラＡ用として１５００と１６００を、コントローラＢ用として２５００、２６００を使用することもできる。この場合、２重書き時と比べ、キャッシュメモリの使用効率は、２倍となる。この１重書き、２重書きの制御はディスクアレイ制御部１２００、２２００が、転送制御部１１００、２１００の転送モードを設定することにより実現される。本キャッシュメモリ１３００、２３００は、ＲＤキャッシュ、及びＷＲキャッシュとして用いられ高いパフォーマンスを提供することが可能である。ＷＲキャッシュ機能を使用するには、キャッシュメモリ障害、及びコントローラ障害時のデータ保証に為、前記記述のように２重書きすることが必須となっている。

【００１６】ホストは、各々２つのポートを持ち、各々別のコントローラへ接続される。図１では、ホスト１０のポートＡ：１１がＳＣＳＩバス１００Ａを介して、コントローラＡ：１０００内ポートＡ：１００Ａへ、ポートＢ：１２－１００Ｂ－ポートＡ：２０１０、以下同様に、別コントローラに接続されている。本接続は、ホスト側の障害、バス障害時、他系に切り替えられる冗長度を持つ構成の代表例である。

【００１７】しかし、本ディスクアレイ装置５０００の複数ポートに、全て別々のホストに接続されてもよい。この場合ホスト側の冗長度はない。又、ホストが、図１のような２バス構成である時、交代系のバスを同一コントローラ内の別ポートに接続してもよい。本構成をとる場合、コントローラ側の障害時、交代バスがなくアクセスが不能となるが、ホスト側、バス障害時には、同一コントローラへの切替えが可能であり、後記述する切替え処理、及び通信による依頼処理どちらも必要でないため、切替えによる性能劣化がないのが特徴である。

【００１８】以下、図１の構成における論理ボリュームの割当てについて説明する。論理ボリュームの割当ては、コントローラ間で排他的に行われる。例えば、コントローラＡ１０００にはＬＵＮ０、１、２、３を、コントローラＢ２０００にはＬＵＮ４、５、６、７を担当に設定する。担当論理ボリュームは、操作パネル３０００で予め設定してもよいし、ディスクアレイ装置の立ち上げ後、受領したコマンドのＬＵＮ番号を担当としてもよい。このように、担当ＬＵＮを分割し、且つコントローラ単位に使用するキャッシュメモリを分割することにより、マルチコントローラ混成特有の排他制御等が不必要となり、高いパフォーマンスを提供することが可能となる。

【００１９】コントローラ内の複数のＬＵＮは、複数のホスト間で共有してもよいし、ホスト毎に独立にアクセスしてもよい。この時、コントローラ内のキャッシュメモリの割当てをコントローラ内で一元的に管理するか、各ＬＵＮ毎に分割して使用するかをユーザが指定することができる。各ホストに独立にＬＵＮを設定した場合、ホスト側のアクセスパターン、ホスト間の能力差等により、任意のホスト配下のＬＵＮのみがキャッシュメモリを占有してしまうと、他ホストからの別ＬＵＮに対するパフォーマンスが低下する。本現象を回避するために、予めキャッシュメモリをＬＵＮ毎に分割することが可能である。本指定も操作パネル３０００により設定することができる。又、ＬＵＮ毎のアクセスの頻度が時間的に変化するような場合、例えば、日中の業務にはＬＵＮ

０，１を夜間ではＬＵＮの２，３をメインにアクセスするような場合には、ＬＵＮ単位の割当てを負荷状態により、ダイナミックに変更することも可能である。ＬＵＮ単位の割当てを静的に配置するか動的に割り当てるかの指定も操作パネルにて設定することができる。本発明により、マルチホスト接続環境での複数論理ボリュームへのパフォーマンスがホストからのいかなるアクセスパターンにおいてもホストへの均等なサービスを提供することができる。

【００２０】次に、ＬＵＮ毎の割当て方法と、負荷状態にあわせての割当て方法の変更方法を図５を使用し説明する。キャッシュメモリの使用／未使用を管理する最小管理単位をセグメントと称す。データ管理情報はＬＵＮ毎に使用可能なセグメント数１４３１と、現在使用中のセグメント数１４３２、未使用のセグメントの数１４３３、各々のセグメント毎のキャッシュの実態ＡＤＲや、ホストからのデータＡＤＲを管理するセグメント管理情報１４３４からなる。セグメント管理情報は、セグメントの数と等しい数の情報を持つ。例えば、コントローラＡ用のキャッシュメモリ１５００の大きさが、１，０００個のセグメント分だとする。今、コントローラＡには、４つのＬＵＮが割り当てられ各々２５０個セグメントに分割指定されたとする。よって、ＬＵＮ毎の使用可能セグメント数１４３１には２５０が設定される。ホストからＩ／Ｏを受領しキャッシュを使用する時、使用中セグメント数１４３２にくわえ、未使用セグメント数１４３３から引き、セグメント管理情報１４３４を設定する。ただし、キャッシュをＲＤキャッシュをして使用する場合は未使用のままで管理し、ＷＲキャッシュとして使用する場合のみ使用中として管理する。

【００２１】また、一定期間で受領したＬＵＮ毎のＩ／Ｏ数をカウントしておく。本情報はディスクアレイ制御部に持ってもよいし、キャッシュメモリの管理情報内に持ってもよい。本情報を元にＬＵＮ毎の負荷を判定し、各ＬＵＮ毎の平均Ｉ／Ｏ数より少ないＬＵＮから多いＬＵＮへ空きセグメントを移行させる。つまり、使用可能セグメント数１４３１の変更と、未使用セグメント数１４３３の変更、及びセグメント管理情報の所在を移行することにより実現される。本変更処理は、一定期間毎に行ってもよいし、ホストからのＩ／Ｏの負荷の低い時に行ってもよい。本制御により、ＬＵＮ毎の負荷によりダイナミックにキャッシュの割当てが可能となる。

【００２２】次に障害時の切替え処理について説明する。図３にＬＵＮの割当ての１例を示す。本割当て例はコントローラ毎にＬＵＮが分割されているだけでなく、ホスト毎、ポート毎にＬＵＮが独立に設定されている。この構成で、ホスト１０のＳＣＳＩバス１００Ａが障害となった時、ホスト側ＬＵＮ０へのアクセスをは交代バス１００Ｂより行う。この時、コントローラＢはＬＵＮ０はコントローラ毎のＬＵＮ情報１４２０と比較し、非担当であることを認識する。この時、コントローラＡ担当分のＬＵＮをすべて一括してコントローラＢに切り替えた場合、残りのホスト２０，３０，４０がポートＡ側からＬＵＮ１、２、３をアクセスする場合の処理もコントローラＢ側の担当に切り替わることになる。したがって、ホスト２０、３０、４０からのＬＵＮ１、２、３に対するアクセスをコントローラＡが受領した場合、非担当ＬＵＮに対するアクセスと判断され、もともとこれらのＬＵＮの担当であったコントローラＡ側にＬＵＮの処理担当を切り替える処理が発生する。このように、コントローラ配下のＬＵＮを一括して切り替える方式では、このようにマルチホスト接続環境では、任意ホスト系障害時、切り替え処理が頻発し、性能が大幅にダウンするといった問題がある。

【００２３】本発明では、前記の問題を解決するため、ＬＵＮ単位の切替えを実現することが可能である。ＬＵＮ単位の切替えを行うと、上記障害時、コントローラＢがＬＵＮ０のＩ／Ｏを受領した時、切替え対象ＬＵＮを当該ＬＵＮ０のみとすることで、他ホストからのコントローラＡへのアクセス時も切り替え処理が発生せず、性能維持が可能である。

【００２４】一方、前記の構成で、コントローラＡが障害となった場合は、全てのホストがポートＢにアクセス権を切り替えてくる為、ＬＵＮごとの切り替えより一括切替えの方が、ＬＵＮの処理担当を切り替える回数が減少し、ＬＵＮ処理担当切り替え処理によるの性能劣化を防ぐことができる。

【００２５】また、図４にしめすような、ＬＵＮの割当て方であった場合、つまり、コントローラ内の各ポート間、つまり、各ホスト間でＬＵＮをシェアして使用する場合、コントローラ側障害時は上記一括切替えにて処理続行が可能であるが、ホスト側障害又は、ＳＣＳＩバス障害時は上記２つの切替え処理でも、切り替え動作が頻繁に発生してしまい、パフォーマンスが大幅にダウンしてしまう。例えば、ホスト１０のＳＣＳＩバス１００Ａが障害となった場合、ホスト１０は、コントローラＡ側での処理を交代バス１００Ｂを使用しコントローラＢに発行してくる。つまり、ＬＵＮ０、１、２、３に対する要求をコントローラＢに発行する。

【００２６】コントローラＢは、非担当ＬＵＮのＩ／Ｏを受領したことを契機に切り替え処理を行う。一括切替えの場合、コントローラＢ側に全てのＬＵＮの担当が移行する。その後、他ホスト２０、３０、４０からＬＵＮ０，１，２，３に対する処理要求がコントローラＡを介してあったとき、再び、一括切替えが発生する。すなわち、ホスト１０からのＩ／Ｏ及びホスト２０、３０、４０からのＩ／Ｏの度に一括切替えが起こる。ＬＵＮ単位切り替えでも、まったく同様な切り替えの繰り返しが起こる。

【００２７】前記の問題を解決する為、本発明は非担当

ＬＵＮに対するＩ／Ｏを受領した時、切り替えない方式も提案している。本方式は、コントローラ間通信情報１４１０を使用し、相手コントローラに相手コントローラエリアのキャッシュの確保とディスク装置４０００との読み出し、書き込みを要求する。相手コントローラは、要求処理が終了すると、コントローラ間通信情報１４１０にその旨を報告する。そして自コントローラは、対ホストとの転送のみを実行する。本発明により、コントローラ内のＬＵＮを複数ホスト間で共有する構成で、ホスト側やバス障害時の交代パスからの処理の続行を頻繁な切替え動作なしに、実行することが可能である。

【００２８】本発明では、一括切り替え処理方式、ＬＵＮ毎の切替え処理方式、切替えなしに相手コントローラへの依頼処理方式の３つのモードから、ユーザが自らのシステム構成のアクセス環境を考慮して最適な方式を選択することができる。この選択は、操作パネル３０００を介して行う。さらに、ユーザがＬＵＮ毎の切り替え処理方式を選択した場合でも、コントローラに障害が発生した場合は、一括切り替え方式によりＬＵＮ処理担当を切り替えるようにしてもよい。

【００２９】又、本発明では、ディスクアレイ装置５０００側でＩ／Ｏパターンより自動選択する自動モードも設定することができる。自動モード時の選択方法を以下に説明する。

【００３０】ＬＵＮの割当て時、コントローラ毎にＬＵＮを割り当てた場合、つまり、各ポート毎に設定しなかった場合は、コントローラ内のＬＵＮを各ポート（ホスト）間でシェアするか独立なのかの判断ができない。よって、一定期間の各ポートからのアクセスＬＵＮの種類をカウントし、一定期間後シェアであるか、独立であるかを判断する。独立であれば、ＬＵＮ単位の切替えを選択するし、シェアであれば、依頼方式を選択する。尚、アクセスパターンが、時間と共に変わる可能性が有る為、さらに一定期間後に上記処理を行う。

【００３１】一括切替えは、シングルポートでのデュアル構成で、交代パス構成のシングルホスト構成時等に有効である。又、図１の様な複数ホスト構成で、ＬＵＮをシェアしてアクセスしている時にも、ホスト間で切り替えたことを通知する。任意ホストの切替え処理に伴い、他ホストも一括して交代パスに切り替える様な制御が可能なホスト環境においては有効である。

【００３２】又、本発明の記憶サブシステムでは、コントローラの各ポート毎に処理を担当するＬＵＮを割り当てるとき、１つのコントローラ内の複数のポートであれば同一のＬＵＮの処理を担当するように割り当てることができる。この場合も、コントローラ内の複数のポートに対するＬＵＮの割り当て方法に基づいて、コントローラがホストコンピュータから処理を担当していないＬＵＮに対する処理要求を受け取ったときに、全ＬＵＮの切替え方式か、ＬＵＮ毎の切替え方式か、依頼方式かを自動的に選択するようにできる。

【００３３】次に、一括切り替え処理の処理手順について説明する。コントローラＢが、非担当ＬＵＮを受領したとする。この時、コントローラ間通信情報１４１０にて、その旨をコントローラＡに伝える。コントローラＡは当該情報により、一括切替えを認識し、現在行っている処理を中断する。中断したら、その旨をコントローラＢに、コントローラ間通信情報１４１０を使用し伝える。コントローラＢは、コントローラＡからの報告を受けると、コントローラ毎のＬＵＮ情報１４２０の変更を行い、コントローラＡ配下の全ＬＵＮの担当をコントローラＢに設定する。又、データ管理情報１４３０のコントローラＡ用の管理情報も、全てコントローラＢ側で管理することになる。この様に、一括切替えの場合は、管理情報の更新のみで切り替えることが可能であり、高速な切替え手段を提供することができる。

【００３４】次にＬＵＮ毎の切替え方法について説明する。コントローラＢ側が、非担当ＬＵＮを受領し、コントローラＡから切り替える場合を説明する。コントローラＢが非担当ＬＵＮを受領したとする。この時、コントローラ間通信情報１４１０にて、その旨をコントローラＡに伝える。コントローラＡは、当該情報により、ＬＵＮ単位切替えを認識し、現在行っている処理を中断する。中断することにより、コントローラＡが、現在キャッシュメモリをアクセスしていないことを保証する。中断したら、その旨をコントローラＢに、コントローラ間通信情報１４１０を使用し伝える。コントローラＢは、コントローラＡからの報告を受けるとコントローラＡ用、Ｂ用のデータ管理情報１４３０を一旦コントローラＢが管理する。コントローラＢは、切り替え対象ＬＵＮの使用中セグメント数１４３２と、当該セグメントの管理情報１４３４を参照し、使用中の数だけ、コントローラＢ側のデータ管理情報内の各ＬＵＮ毎の未使用セグメント数と、当該セグメント管理情報より、コントローラＡの使用中であるセグメント内データをコントローラＢ側の空きセグメントへコピーする。本コピー処理自体は、コントローラＢ内の転送制御部２１００により実現することができる。

【００３５】データ自体のコピーと共に、コントローラＡ側の切替え対象ＬＵＮのＬＵＮ毎のデータ管理情報を削除し、元々、当該ＬＵＮに使用可能セグメント数として割り当てられていたセグメントを他のＬＵＮに分配する。分配されたセグメントは、未使用セグメント数として登録される。又、反対にコントローラＢ側のデータ管理情報は、切り替え対象ＬＵＮ情報を追加する。当該ＬＵＮに、コピーした数分使用中セグメントとして登録し、使用可能セグメント数は、他ＬＵＮ内の使用可能セグメント数から一部ずつ移行させ、空きセグメントとして登録する。他ＬＵＮの使用可能セグメント数も一部ずつ減少させ、また、コピーする為に割り当てた数の空き

セグメントも空きセグメント数から減少させる。本一連の処理が完了したところで、コントローラＢは、コントローラＡに処理の再開をコントローラ間通信情報１４１０を使用し伝える。

【００３６】また、本制御を実現するために、キャッシュ内で使用中セグメントの数、つまり、ＷＲキャッシュ機能により、ディスク装置４０００に未反映のＷＲデータの数を管理し、一定量以上に達した場合は、ディスク装置への書き込み処理を最優先に行うことにより、一定量以上に増加しないよう管理する。本制御は、ディスクアレイ制御部１２００、２２００がＬＵＮ毎の使用中セグメント数を一定周期で参照し、一定量以上に達したと判断すると、前記書き込み処理を優先的に処理することにより実現される。本制御により、常に一定量以上の空きセグメントが、キャッシュ内に存在することにより、ＬＵＮ単位のコピー方式での切り替えが可能となる。本切替え処理は、キャッシュ間でＷＲデータのコピー処理が必要となるのでキャッシュ搭載容量、また、その時のＷＲデータ量により、処理時間が増加する。しかし乍ら、近年、データバスの能力は非常に高くなっており、キャッシュ間のコピー処理も数１００ＭＢ／Ｓの能力を持つものが一般的であるため、本切り替え処理時間も数秒程度で完了するため、ホストとの整合性上問題はない。

【００３７】

【発明の効果】本発明によれば、コントローラ当たりホストとのバスプロトコル制御を行う複数のポートを持ち、コントローラ及びキャッシュを２重化した記憶サブシステムにおいて、複数ホスト間との接続において、コントローラ毎に独立に論理ボリュームの割当てを行うことにより、いかなる接続形態においても、通常時、複数コントローラ制御による排他制御等が必要なくなり、シングル構成と比べ、高いパフォーマンスが提供できる。又、ホスト側障害、接続バス障害、コントローラ側ポート障害時、ホストが、交代パスを同一コントローラ内の別パスに接続している場合には、切替え処理が必要なく、性能劣化のない交代処理が可能となる。又、別コントローラに交代パスを接続している場合には、コントローラ側障害時にも冗長度があるため、全ての障害において、切り替え処理が可能となる。又、ホストからのＬＵＮへのアクセス形態により、切替え時間の少ない最適な切替え手段を提供することができる。

【図面の簡単な説明】

【図１】本発明の実施例である制御装置の構成図である

【図２】本発明の実施例であるコントローラのキャッシュ内のデータの構成図である。

【図３】本発明の実施例である複数ホスト間との論理ボリュームの割当て方の一例であり、ホスト毎、ポートごとに独立した論理ボリュームを割り当てている例である。

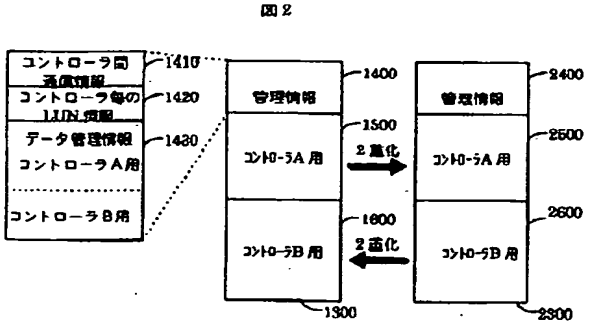
【図４】本発明の実施例である複数ホスト間との論理ボリュームの割当て方の一例であり、ホスト間で論理ボリュームを共有している一例である。

【図５】本発明の実施例によるキャッシュメモリ内の管理情報の構成図である。

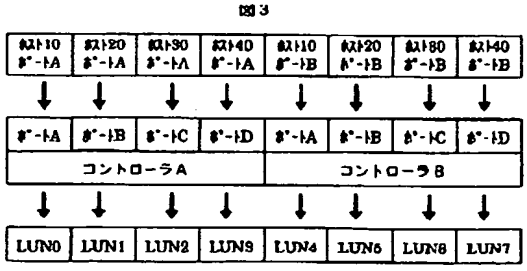
【符号の説明】

１０、２０、３０、４０：ホストコンピュータ
１１、２１、３１、４１：ポートＡ
１２、２２、３２、４２：ポートＢ
１００Ａ～４００Ａ：ＳＣＳＩバスＡ
１００Ｂ～４００Ｂ：ＳＣＳＩバスＢ
１０１０～１０４０、２０１０～２０４０：ポートＡ～Ｄ
１１００、２１００：転送制御部
１２００、２２００：ディスクアレイ制御部
１３００、２３００：キャッシュメモリ
１０００、２０００：コントローラ
３０００：操作パネル
４０００：ディスク装置群
５０００：ディスクアレイ装置

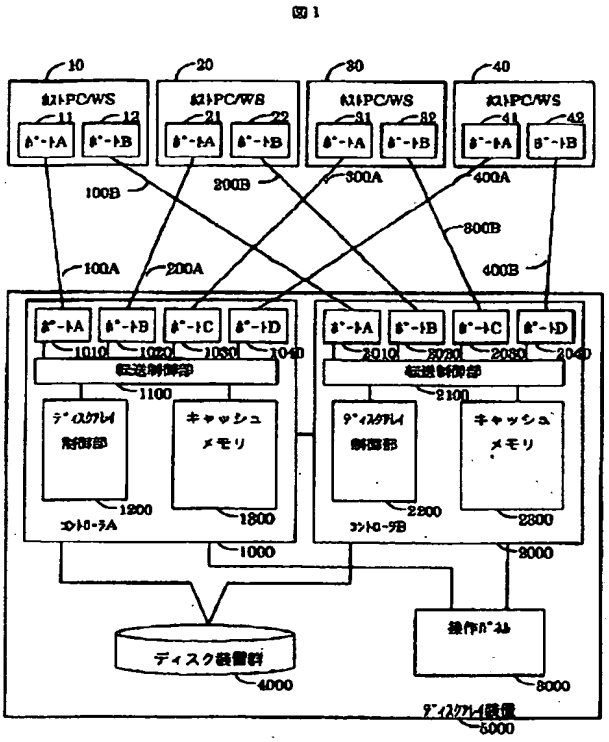
【図２】



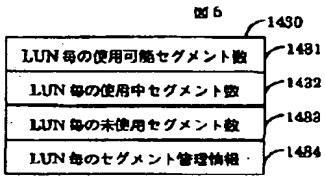
【図３】



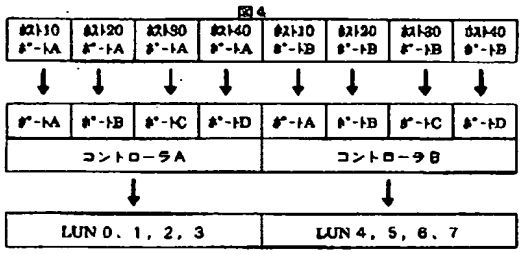
【図１】



【図５】



【図４】



フロントページの続き

(72) 発明者　小林　正明
神奈川県小田原市国府津2880番地　株式会
社日立製作所ストレージシステム事業部内

【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第６部門第３区分
【発行日】平成17年9月29日(2005.9.29)

【公開番号】特開平11-312058
【公開日】平成11年11月9日(1999.11.9)
【出願番号】特願平10-118127
【国際特許分類第７版】
　Ｇ０６Ｆ　3/06
　Ｇ０６Ｆ　11/20
　Ｇ０６Ｆ　12/08
　Ｇ０６Ｆ　12/16
【ＦＩ】
　Ｇ０６Ｆ　3/06　３０４Ｂ
　Ｇ０６Ｆ　11/20　３１０Ｆ
　Ｇ０６Ｆ　12/08　Ｇ
　Ｇ０６Ｆ　12/08　Ｊ
　Ｇ０６Ｆ　12/16　３１０Ｊ

【手続補正書】
【提出日】平成17年4月18日(2005.4.18)
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】特許請求の範囲
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
　複数の論理ボリュームに分割された記憶領域を有する記憶装置群と、ホストコンピュータに接続され、前記ホストコンピュータと前記記憶装置群との間を転送されるデータを一時的に保持するキャッシュメモリを有する複数のコントローラと、を備え、前記複数のコントローラが前記記憶装置群を制御する記憶サブシステムであって、
　前記コントローラは、処理を担当する前記論理ボリュームが割り当てられ、前記キャッシュメモリは、前記複数のコントローラのための記憶領域を有することを特徴とする記憶サブシステム。
【請求項２】
　請求項１に記載の記憶サブシステムであって、
　前記コントローラは、前記複数のコントローラのための前記キャッシュメモリの記憶領域を、前記コントローラが処理を担当する論理ボリュームのための領域に分割することを特徴とする記憶サブシステム。
【請求項３】
　請求項１に記載の記憶サブシステムであって、
　前記コントローラが前記ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領したときは、前記コントローラは、前記論理ボリュームの処理を担当していたコントローラが担当する全ての論理ボリュームの処理を担当することを特徴とする記憶サブシステム。
【請求項４】
　請求項１又は請求項２に記載の記憶サブシステムであって、
　前記コントローラが前記ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領したときは、前記コントローラは、前記論理ボリュームの処理を担当する

ことを特徴とする記憶サブシステム。
【請求項５】
請求項１に記載の記憶サブシステムであって、
前記コントローラは、前記ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領したときは、前記論理ボリュームの処理担当コントローラに処理要求内容を通信し、通信を受領したコントローラが前記論理ボリュームに対する処理を行い、処理結果を要求元のコントローラに通信することを特徴とする記憶サブシステム。
【請求項６】
請求項１に記載の記憶サブシステムであって、
前記コントローラが前記ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受信したときの処理方法を選択するための入力手段を有することを特徴とする記憶サブシステム。
【請求項７】
請求項２に記載の記憶サブシステムであって、
コントローラ当たりのキャッシュメモリ領域の容量、及び前記論理ボリュームのためのキャッシュメモリ領域の容量を指定するための入力手段を有することを特徴とする記憶サブシステム。
【請求項８】
請求項１に記載の記憶サブシステムであって、
前記複数のコントローラのキャッシュメモリのそれぞれに前記複数の論理ボリュームのデータを格納することを特徴とする記憶サブシステム。
【請求項９】
請求項１に記載の記憶サブシステムであって、
前記複数のコントローラのキャッシュメモリのそれぞれに前記複数の論理ボリュームのデータを格納するか、或いは前記コントローラのそれぞれが処理を担当する論理ボリュームのデータのみを格納するかを選択する入力手段を有することを特徴とする記憶サブシステム。
【請求項１０】
請求項１に記載の記憶サブシステムであって、
前記コントローラに障害が発生した場合には、障害コントローラが担当していた論理ボリュームの処理を正常に動作しているコントローラの担当に切り替えることを特徴とする記憶サブシステム。
【請求項１１】
請求項３に記載の記憶サブシステムであって、
前記コントローラが前記ホストコンピュータから処理を担当していない論理ボリュームに対する処理要求を受領したときは、キャッシュメモリ上の管理情報のみの変更で担当の切り替えを行うことを特徴とする記憶サブシステム。
【請求項１２】
請求項４に記載の記憶サブシステムであって、
前記コントローラが前記ホストコンピュータから処理を担当していない論理ボリュームに対する処理要求を受領したときは、キャッシュメモリ上の切り替え対象論理ボリュームのうち前記記憶装置群の記憶領域に格納されていないデータを切り替え先のコントローラ内のキャッシュメモリ上にコピーして前記論理ボリュームの処理を担当するコントローラを切り替えることを特徴とする記憶サブシステム。
【請求項１３】
請求項２に記載の記憶サブシステムであって、
前記コントローラは、前記キャッシュメモリの論理ボリュームのための記憶領域を、前記論理ボリュームへ入出力されるデータ量に応じて変更することを特徴とする記憶サブシステム。
【請求項１４】

複数の論理ボリュームに分割された記憶領域を有する記憶装置群と、ホストコンピュータに接続する複数のポートを有する複数のコントローラとを備える記憶サブシステムであって、

各コントローラは、処理を担当する論理ボリュームがポート毎に排他的に割り当てられており、

所定のコントローラの所定のポートに接続するアクセスパスに障害が生じると、障害が生じたポートに割り当てられた論理ボリュームの処理を、正常に動作しているコントローラの担当に切り替えることを特徴とする記憶サブシステム。

【請求項１５】

複数の論理ボリュームに分割された記憶領域を有する記憶装置群と、ホストコンピュータに接続する複数のポートを有する複数のコントローラとを備える記憶サブシステムであって、

各コントローラは、処理を担当する論理ボリュームがポート毎に排他的に割り当てられており、

所定のコントローラに障害が生じると、障害が生じたコントローラに割り当てられていた全ての論理ボリュームの処理を、正常に動作しているコントローラの担当に切り替えることを特徴とする記憶サブシステム。